DOCUMENT RESUME

ED 448 210 TM 032 240

AUTHOR McDonald, Jo-Anne
TITLE The Susceptibility of Item Parameters to Instructions for Completion.
PUB DATE 2000-04-00
NOTE 19p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Employees; Item Response Theory; *Personality Measures; Rating Scales; Responses; *Selection
IDENTIFIERS *Instructions; *Item Parameters; Rasch Model; Self Report Measures

ABSTRACT

The purpose of this study was to examine the effect of completion instructions on item parameters and category use as students completed a self-report personality survey. Instructions allowed free-choice allotment of ratings (nonforced distribution) or requested the subject to assign a certain number of ratings to either the highest or lowest rating-scale categories (forced instructions). A total of 126 participant responses using nonforced instructions was obtained from a self-report survey completed at 1 company. The comparison sample of 346 forced distributions of ratings was collected from 5 companies. The hypothesis that alternatively worded instructions for survey completion would not impact item parameters was unsupported. The two sets of instructions produced nonequivalent patterns of response and item statistics. In addition, person separation and reliability were lower in the forced distribution condition. Data collected under the forced conditions did not fit the Rasch model as well as did the nonforced distribution data. Five appendixes contain the instructions for both conditions, two figures, and three tables of data. (SLD)

The Susceptibility of Item Parameters to Instructions for Completion.

Paper presented at AERA
New Orleans, 2000
Jo-Anne McDonald, MS
University of Denver
Denver, Colorado

The susceptibility of item parameters to instructions for completion.

Introduction

"A statement about an empirical system is meaningful only when it is scale independent, that is, only when it is true on all of the permissible numeric scales".

-Townsend and Ashby (1984, pp. 399)

According to Wright and Stone (1979), one of the characteristics of Rasch-based measurement is that the difficulty of an item is independent of a person's degree of endorsement of it. The item measure is said to be invariant across groups and individuals. In other words, the item is difficult relative to other items and not related to the magnitude of a person's response to it. As well, the theory infers that the item difficulty will remain the same no matter what scale we use to measure it by (Englehard, 1992).

However, few studies have been done to assess the impact of instructions on results of Rasch analyses. There is a dearth of evidence in the literature on equivalence of item parameters in studies involving manipulations of response formats and instruction sets.

The purpose of the present investigation was to examine the effect of completion instructions on item parameters and category usage. Instructions a) allowed free-choice allotment of ratings (non-forced instructions) and b) requested the subject to assign a certain number of ratings to either the highest or lowest rating-scale categories (forced instructions). It was hypothesized that:

H_1: There would not be significant effects of instructions on the item calibrations.
H_2: There would be no effects of instructions on the person parameter separation or reliability.

Method

Subjects.

A total of 126 participants using the non-forced distributions (NFD) were obtained from a self-report survey completed in 1996 in one company. The comparison sample of 346 forced distributions (FD) of ratings was collected from 5 companies between 1996 and 1998. Cases were deleted for recording errors and missing responses. Samples of equivalent size (n = 115), that contained complete sets of data on the 6-item Group Involvement scale, were selected from the two instructional-set groups for purposes of comparison.

Procedure

Copies of the two forms of the personality survey had been distributed and collected in previous investigations in companies across the United States. The 1996a version (NFD) had been administered to employees of a large utility company in 1996. The 1996b version (FD) had been used to survey employees in 5 large public and private corporations between 1996 and 1998. Responses were returned directly to the researchers and individual responses were kept confidential. Data from the surveys were analyzed and compared across instruction groups using the WINSTEPS (Linacre, 1998) program for Rasch analysis.

Instrument.

The instrument used to collect these data was the Employee Profile© (Somerville and Company, Inc. 1996a, 1996b), a 144-item survey of personality characteristics relevant to work-related contexts. The 1996a version (NFD) contained directions to the respondents to rate themselves on a scale of 1 to 6 points according to how well the statements described their modes of behavior in work-related situations. In addition, the instructions requested that subjects distribute their ratings across the entire 6-point scale. See Appendix A for a sample item and its response format.

The 1996b version (FD) asked participants to rate themselves on a 10-point scale (1 to 10). As well, they were to put at least 20 of the ratings into either of the two lowest categories (1 and 2) and to not put more than 20 of the ratings into either of the two highest categories (9 and 10). A sample item with this format is included in Appendix B.

The Group Involvement scale consists of 6 items that describe work-related behaviors. Group involvement describes the propensity toward involving one's self in team efforts and to publicly recognize and promote members' contributions. See Appendix C for the items.

Results

It was hypothesized that item difficulties would differ across instruction groups. To test this hypothesis a Rasch model analysis was performed on each set of data. A t-value was computed to test for differences between groups on each item. The reported t-value was calculated according to Wright and Stone (pp. 95,1979) for the degree to which the sample difficulty measures approximate the same difficulty parameter. In Table 1 it may be observed that 4 of the 6 parameter estimates differed significantly ($p < .05$). Also from Table 1, items 2 and 6 exchanged rank order. The difference between the Item 2 and Item 6 parameters was 0.24 logits ($t = 1.41$, n.s.) in the NFD sample and 0.06 logits (t = .004, n.s.) in the FD sample.

Table 1.
Comparison of item parameter estimates across instructional sets.

Item	Set	Measure	Error	t-value	Rank
1	NFD	-.45	.13	-2.51	5
	FD	-.10	.05		5
2	NFD	.31	.12	2.15	2
	FD	.03	.05		4
3	NFD	.54	.11	3.08	1
	FD	.18	.04		1
4	NFD	.21	.12	1.03	3
	FD	.08	.04		3
5	NFD	-.69	.13	-2.86	6
	FD	-.28	.06		6
6	NFD	.07	.12	0.16	4
	FD	.09	.04		2

Apparently, the 6 items represented an underlying, unidimensional construct. None of the standardized fit statistics were greater than 2.0. However, it was interesting to note that the point biserial correlations of the items with the full scale were consistently higher in the NFD sample. Correlations for each sample significantly differed on all but Items 3 and 5 (two-tail test, $p < .05$).

Examination of results presented in Table 3 revealed that the person ability parameter was measured more reliably using the NFD rather than the FD instruction set. Real and model reliability estimates significantly differed between the groups (two-tail test, $p < .05$). The NFD sample displayed greater separation between persons than did the FD sample.

Table 2.
Comparison of item fit statistics across instructional groups.

Item	Set	Infit MNSQ	z MNSQ	Outfit MNSQ	z MNSQ	Point biserial
1	NFD	.87	-.9	.87	-1.0	.52
	FD	.95	-.4	.93	-.4	.24
2	NFD	.99	-.1	1.05	.4	.49
	FD	1.06	.5	1.07	.5	.22
3	NFD	1.03	.2	.98	-.1	.50
	FD	.95	-.5	.88	-1.0	.30
4	NFD	1.04	.3	1.01	.1	.51
	FD	.88	-1.1	.90	-.8	.24
5	NFD	1.04	.3	1.05	.3	.46
	FD	1.04	.2	.97	-.2	:32
6	NFD	.95	-.4	.97	-.2	.47
	FD	1.17	1.5	1.22	1.6	.04

Table 3.
Summary of measured persons.

		RSME	Adjusted Std. Dev.	Separation	Reliability
NFD	Real	.60	1.00	1.68	.74
	Model	.53	1.04	1.95	.79
FD	Real	.26	.27	1.06	.53
	Model	.23	.30	1.29	.62

As has been seen in Table 1, the item measures had greater variability in the NFD than the FD condition (Table 4). As a result, there was a wider range of item separation under the NFD set of instructions.

Table 4.
Summary of measured items.

		RSME	Adjusted Std. Dev.	Separation	Reliability
NFD	Real	.12	.41	3.40	.92
	Model	.12	.41	3.43	.92
FD	Real	.05	.14	2.93	.90
	Model	.05	.14	3.0	.90

Discussion

The principle of invariance of item parameters suggests that the scales used to measure attributes should not affect the difficulty or relative order of the same items measured two different ways. The sample values of the item difficulty measures are expected to differ somewhat but the order and separation of the items would be expected to remain constant across variations of rating scale.

However, the hypothesis, that alternatively worded instructions for survey completion would not impact item parameters, was unsupported. Asking participants to use the full range of their response options versus asking them to assign a proportion of their responses to the extreme categories produced non-equivalent patterns of response and item statistics. In addition, person separation and reliability were lower in the FD condition. It may be concluded that the data collected under the FD instructions did not fit the Rasch model as well as did the NFD data.

It should be noted that the efficiency of the rating scales differed between the samples, as well. The utilization of the scale categories was not the same across instruction sets and rating formats. See Appendices D and E. It may be that imposing limits on responses in certain categories, as was the case with the FD instructions, on a rating scale provoked categorical responses like "this is one of my 20 best/worst characteristics – yes or no".

As well, Andrich (1996) and van der Linden (1993) have described conditions under which the number of categories on rating scales impacted the fit of the data to the model. However, collapsing the categories may not accurately reflect the true responses or abilities of the persons on the items. Further investigations should include the reduced scale formats for empirical conformation of category utilization.

The contribution of the present investigation to knowledge in the field of measurement and survey design is twofold. As Townsend and Ashby (1984) pointed out, studies of alternative rating methods are needed before we can assume findings generalize to all applications. The results serve as a cautionary note to survey designers and consumers. In this case, forcing categorical decisions onto a rating scale along with decisions that were to be made on a continuum created rather messy distributions. And it appeared to have an impact on statistics derived from the sample obtained under these conditions. Certainty in conclusions based on sample data may be compromised by violations of the assumption that item difficulty remains invariant across types of instructions and rating formats.

And the results were informative in that two methods of distributing self-report ratings, one of which appears to have remained unreported in the literature, were analyzed and compared using Rasch model analysis. Survey researchers face tremendous odds against finding high-precision methods of measurement in surveys of characteristics, attitudes, and behaviors. Rasch modeling provides survey researchers a useful avenue of investigation and basis for comparison of alternative techniques. Reports of investigations into the impact of instructional sets on responses help to direct the search for more accurate ways to measure people.

References

Andrich, D. (1996). Category ordering and their utility. Rasch Measurement Transactions, 9, 464-465.

Englehard, G. (1992). Historical views of invariance: Evidence from the measurement theories of Thorndike, Thurstone, and Rasch. Educational and Psychological Measurement, 52, 275-291.

Townsend, J., & Ashby, F. (1984). Measurement scales and statistics: The misconception misconceived. Psychological Bulletin, 96, 394-401.

van der Linden, W. (1993). Measurement scale in the Rasch model. Rasch Measurement Transactions, 7, 287.

Wright, B., & Stone, M. (1979). Best test design. Mesa Press: Chicago.

APPENDIX A

Non-forced distribution instructions to respondents on the Employee Profile.

The Employee Profile Survey was designed to assist in identifying an individual's most and least prominent work behaviors and characteristics. Your ratings should be based on how frequently these behaviors occur or how characteristic that behavior is of you.

Each person should exhibit some traits that are obvious to those around them. The behaviors demonstrated most frequently should be rated 5 or 6. This will identify your most prominent behavioral features.

It would be likely that a person would not demonstrate some of these behaviors very often. Rate some of the items 1 or 2 to identify your least prominent behavioral characteristics.

Respond to the remaining items with 3 or 4 ratings depending upon how often or how characteristic you think these behaviors are of you.

Almost Never ① Not at All Characteristic	Seldom or Once in a while ② Slightly	Regularly but Not often. ③ Moderately Characteristic	Fairly Often ④ Characteristic	Very Frequently ⑤ Very Characteristic	Almost Always ⑥ Extremely

1A. Shows a sincere interest in suggestions from members of the work group. ①②③④⑤⑥

APPENDIX B

Forced distribution instructions to respondents on the Employee Profile.

Read each item and determine how characteristic of you the behavior associated with the item is using the "1" to "10" scale. In order to ensure that your most and least characteristic traits are identified, rate at least twenty (20) items "1" or "2", and no more than twenty (20) items at "9 or 10." You may find that putting 5 items in the "1" or "2" and 5 in the "9" or "10" ranges on EACH page is easiest way to accomplish this. Rate the remaining items within the range "3" and "8" depending upon your assessment of how characteristic of you each item is.

Less Characteristic	①②	③④⑤⑥⑦⑧	⑨⑩	More Characteristic

1B Shows a sincere interest in suggestions from members of the work group ①② ③④⑤⑥⑦⑧ ⑨⑩

APPENDIX C

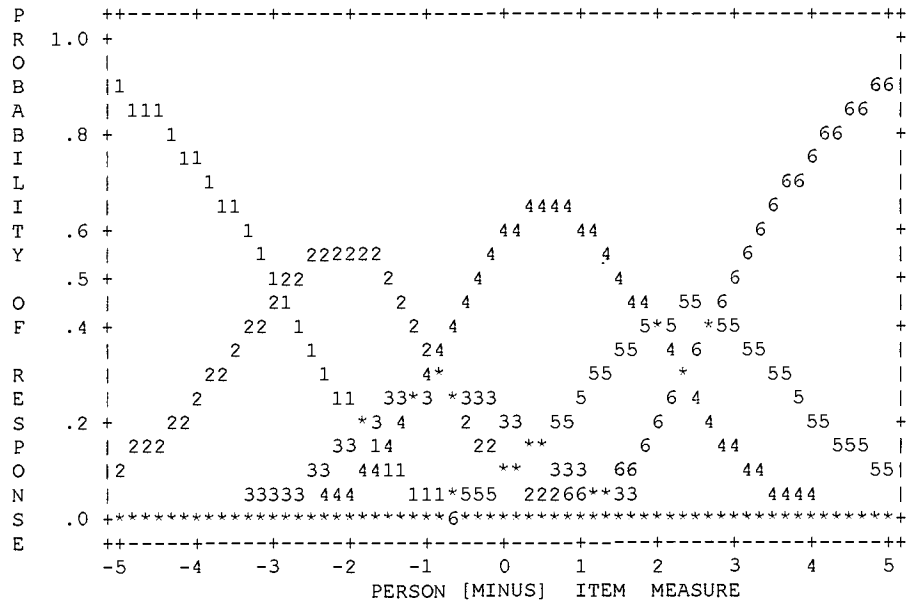
Items on the Group Involvement Scale.

ITEM	STATEMENT
1	Shows a sincere interest in suggestions from members of the work group.
2	Seeks out the special talents and abilities of others to contribute to the quality of the team's product.
3	Encourages people to speak up even when their opinions differ from the opinions of the majority.
4	Develops a feeling of unity and sharing among co-workers.
5	Publicly shares credit for success with those who contributed.
6	Actively promotes the involvement of people having a stake in the outcome of the project or task.

APPENDIX D

Non-forced Distribution Instructions (NFD)

Category probabilities.



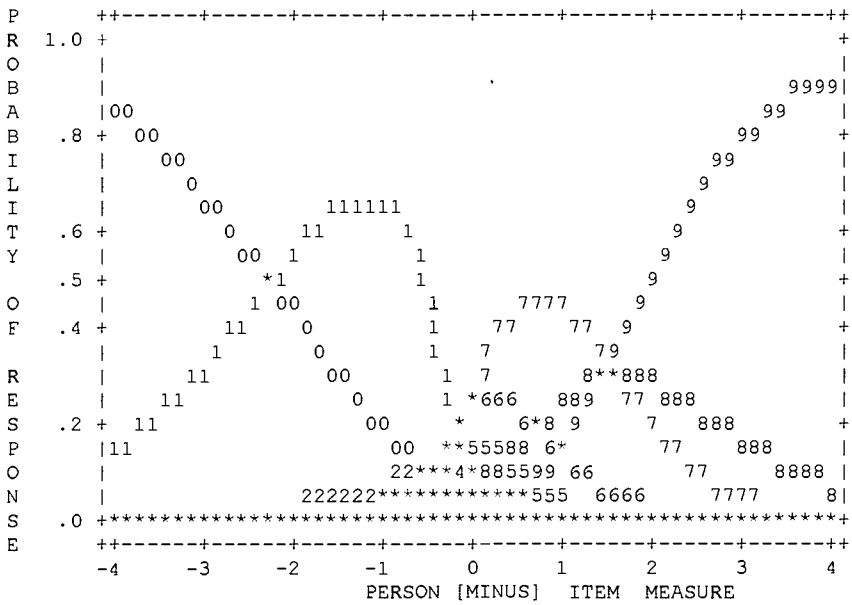
Summary of measured categories.

Category	Number of Observations	Step Calibration	Step Std. Error	Thurstone Threshold
1	7			
2	56	-2.91	.39	-3.01
3	82	-.71	.15	-1.20
4	324	-1.07*	.11	-.52
5	143	1.92	.10	1.69
6	72	2.77	.15	3.05

APPENDIX E

Forced Distribution Instructions (FD)

Category probabilities



Summary of measured categories.

Category	Number of Observations	Step Calibration	Step Std. Error	Thurstone Threshold
1	14			
2	102	-2.25	.27	-2.29
3	25	1.20	.11	-.43
4	21	.03*	.11	-.35
5	48	-.91*	.10	-.30
6	85	-.58	.09	-.21
7	137	-.40	.09	-.06
8	185	-.12	.09	.27
9	52	1.59	.13	1.26
10	21	1.44*	.23	2.00

The Susceptibility of Item Parameters to Instructions for Completion.

Paper presented at AERA
New Orleans, 2000
Jo-Anne McDonald, MS
University of Denver
Denver, Colorado

The susceptibility of item parameters to instructions for completion.

Introduction

"A statement about an empirical system is meaningful only when it is scale independent, that is, only when it is true on all of the permissible numeric scales".

-Townsend and Ashby (1984, pp. 399)

According to Wright and Stone (1979), one of the characteristics of Rasch-based measurement is that the difficulty of an item is independent of a person's degree of endorsement of it. The item measure is said to be invariant across groups and individuals. In other words, the item is difficult relative to other items and not related to the magnitude of a person's response to it. As well, the theory infers that the item difficulty will remain the same no matter what scale we use to measure it by (Englehard, 1992).

However, few studies have been done to assess the impact of instructions on results of Rasch analyses. There is a dearth of evidence in the literature on equivalence of item parameters in studies involving manipulations of response formats and instruction sets.

The purpose of the present investigation was to examine the effect of completion instructions on item parameters and category usage. Instructions a) allowed free-choice allotment of ratings (non-forced instructions) and b) requested the subject to assign a certain number of ratings to either the highest or lowest rating-scale categories (forced instructions). It was hypothesized that:

H_1: There would not be significant effects of instructions on the item calibrations.
H_2: There would be no effects of instructions on the person parameter separation or reliability.

Method

Subjects.

A total of 126 participants using the non-forced distributions (NFD) were obtained from a self-report survey completed in 1996 in one company. The comparison sample of 346 forced distributions (FD) of ratings was collected from 5 companies between 1996 and 1998. Cases were deleted for recording errors and missing responses. Samples of equivalent size (n = 115), that contained complete sets of data on the 6-item Group Involvement scale, were selected from the two instructional-set groups for purposes of comparison.

Procedure

Copies of the two forms of the personality survey had been distributed and collected in previous investigations in companies across the United States. The 1996a version (NFD) had been administered to employees of a large utility company in 1996. The 1996b version (FD) had been used to survey employees in 5 large public and private corporations between 1996 and 1998. Responses were returned directly to the researchers and individual responses were kept confidential. Data from the surveys were analyzed and compared across instruction groups using the WINSTEPS (Linacre, 1998) program for Rasch analysis.

Instrument.

The instrument used to collect these data was the Employee Profile© (Somerville and Company, Inc. 1996a, 1996b), a 144-item survey of personality characteristics relevant to work-related contexts. The 1996a version (NFD) contained directions to the respondents to rate themselves on a scale of 1 to 6 points according to how well the statements described their modes of behavior in work-related situations. In addition, the instructions requested that subjects distribute their ratings across the entire 6-point scale. See Appendix A for a sample item and its response format.

The 1996b version (FD) asked participants to rate themselves on a 10-point scale (1 to 10). As well, they were to put at least 20 of the ratings into either of the two lowest categories (1 and 2) and to not put more than 20 of the ratings into either of the two highest categories (9 and 10). A sample item with this format is included in Appendix B.

The Group Involvement scale consists of 6 items that describe work-related behaviors. Group involvement describes the propensity toward involving one's self in team efforts and to publicly recognize and promote members' contributions. See Appendix C for the items.

Results

It was hypothesized that item difficulties would differ across instruction groups. To test this hypothesis a Rasch model analysis was performed on each set of data. A t-value was computed to test for differences between groups on each item. The reported t-value was calculated according to Wright and Stone (pp. 95,1979) for the degree to which the sample difficulty measures approximate the same difficulty parameter. In Table 1 it may be observed that 4 of the 6 parameter estimates differed significantly ($p < .05$). Also from Table 1, items 2 and 6 exchanged rank order. The difference between the Item 2 and Item 6 parameters was 0.24 logits ($t = 1.41$, n.s.) in the NFD sample and 0.06 logits (t = .004, n.s.) in the FD sample.

Table 1.
Comparison of item parameter estimates across instructional sets.

Item	Set	Measure	Error	t-value	Rank
1	NFD	-.45	.13	-2.51	5
	FD	-.10	.05		5
2	NFD	.31	.12	2.15	2
	FD	.03	.05		4
3	NFD	.54	.11	3.08	1
	FD	.18	.04		1
4	NFD	.21	.12	1.03	3
	FD	.08	.04		3
5	NFD	-.69	.13	-2.86	6
	FD	-.28	.06		6
6	NFD	.07	.12	0.16	4
	FD	.09	.04		2

Apparently, the 6 items represented an underlying, unidimensional construct. None of the standardized fit statistics were greater than 2.0. However, it was interesting to note that the point biserial correlations of the items with the full scale were consistently higher in the NFD sample. Correlations for each sample significantly differed on all but Items 3 and 5 (two-tail test, $p < .05$).

Examination of results presented in Table 3 revealed that the person ability parameter was measured more reliably using the NFD rather than the FD instruction set. Real and model reliability estimates significantly differed between the groups (two-tail test, $p < .05$). The NFD sample displayed greater separation between persons than did the FD sample.

Table 2.
Comparison of item fit statistics across instructional groups.

Item	Set	Infit MNSQ	z MNSQ	Outfit MNSQ	z MNSQ	Point biserial
1	NFD	.87	-.9	.87	-1.0	.52
	FD	.95	-.4	.93	-.4	.24
2	NFD	.99	-.1	1.05	.4	.49
	FD	1.06	.5	1.07	.5	.22
3	NFD	1.03	.2	.98	-.1	.50
	FD	.95	-.5	.88	-1.0	.30
4	NFD	1.04	.3	1.01	.1	.51
	FD	.88	-1.1	.90	-.8	.24
5	NFD	1.04	.3	1.05	.3	.46
	FD	1.04	.2	.97	-.2	.32
6	NFD	.95	-.4	.97	-.2	.47
	FD	1.17	1.5	1.22	1.6	.04

Table 3.
Summary of measured persons.

		RSME	Adjusted Std. Dev.	Separation	Reliability
NFD	Real	.60	1.00	1.68	.74
	Model	.53	1.04	1.95	.79
FD	Real	.26	.27	1.06	.53
	Model	.23	.30	1.29	.62

As has been seen in Table 1, the item measures had greater variability in the NFD than the FD condition (Table 4). As a result, there was a wider range of item separation under the NFD set of instructions.

Table 4.
Summary of measured items.

		RSME	Adjusted Std. Dev.	Separation	Reliability
NFD	Real	.12	.41	3.40	.92
	Model	.12	.41	3.43	.92
FD	Real	.05	.14	2.93	.90
	Model	.05	.14	3.0	.90

Discussion

The principle of invariance of item parameters suggests that the scales used to measure attributes should not affect the difficulty or relative order of the same items measured two different ways. The sample values of the item difficulty measures are expected to differ somewhat but the order and separation of the items would be expected to remain constant across variations of rating scale.

However, the hypothesis, that alternatively worded instructions for survey completion would not impact item parameters, was unsupported. Asking participants to use the full range of their response options versus asking them to assign a proportion of their responses to the extreme categories produced non-equivalent patterns of response and item statistics. In addition, person separation and reliability were lower in the FD condition. It may be concluded that the data collected under the FD instructions did not fit the Rasch model as well as did the NFD data.

It should be noted that the efficiency of the rating scales differed between the samples, as well. The utilization of the scale categories was not the same across instruction sets and rating formats. See Appendices D and E. It may be that imposing limits on responses in certain categories, as was the case with the FD instructions, on a rating scale provoked categorical responses like "this is one of my 20 best/worst characteristics – yes or no".

As well, Andrich (1996) and van der Linden (1993) have described conditions under which the number of categories on rating scales impacted the fit of the data to the model. However, collapsing the categories may not accurately reflect the true responses or abilities of the persons on the items. Further investigations should include the reduced scale formats for empirical conformation of category utilization.

The contribution of the present investigation to knowledge in the field of measurement and survey design is twofold. As Townsend and Ashby (1984) pointed out, studies of alternative rating methods are needed before we can assume findings generalize to all applications. The results serve as a cautionary note to survey designers and consumers. In this case, forcing categorical decisions onto a rating scale along with decisions that were to be made on a continuum created rather messy distributions. And it appeared to have an impact on statistics derived from the sample obtained under these conditions. Certainty in conclusions based on sample data may be compromised by violations of the assumption that item difficulty remains invariant across types of instructions and rating formats.

And the results were informative in that two methods of distributing self-report ratings, one of which appears to have remained unreported in the literature, were analyzed and compared using Rasch model analysis. Survey researchers face tremendous odds against finding high-precision methods of measurement in surveys of characteristics, attitudes, and behaviors. Rasch modeling provides survey researchers a useful avenue of investigation and basis for comparison of alternative techniques. Reports of investigations into the impact of instructional sets on responses help to direct the search for more accurate ways to measure people.

References

Andrich, D. (1996). Category ordering and their utility. Rasch Measurement Transactions, 9, 464-465.

Englehard, G. (1992). Historical views of invariance: Evidence from the measurement theories of Thorndike, Thurstone, and Rasch. Educational and Psychological Measurement, 52, 275-291.

Townsend, J., & Ashby, F. (1984). Measurement scales and statistics: The misconception misconceived. Psychological Bulletin, 96, 394-401.

van der Linden, W. (1993). Measurement scale in the Rasch model. Rasch Measurement Transactions, 7, 287.

Wright, B., & Stone, M. (1979). Best test design. Mesa Press: Chicago.

APPENDIX A

Non-forced distribution instructions to respondents on the Employee Profile.

The Employee Profile Survey was designed to assist in identifying an individual's most and least prominent work behaviors and characteristics. Your ratings should be based on how frequently these behaviors occur or how characteristic that behavior is of you.

Each person should exhibit some traits that are obvious to those around them. The behaviors demonstrated most frequently should be rated 5 or 6. This will identify your most prominent behavioral features.

It would be likely that a person would not demonstrate some of these behaviors very often. Rate some of the items 1 or 2 to identify your least prominent behavioral characteristics.

Respond to the remaining items with 3 or 4 ratings depending upon how often or how characteristic you think these behaviors are of you.

Almost Never ① Not at All Characteristic	Seldom or Once in a while ② Slightly	Regularly but Not often. ③ Moderately Characteristic	Fairly Often ④ Characteristic	Very Frequently ⑤ Very Characteristic	Almost Always ⑥ Extremely

1A. Shows a sincere interest in suggestions from members of the work group. ①②③④⑤⑥

APPENDIX B

Forced distribution instructions to respondents on the Employee Profile.

Read each item and determine how characteristic of you the behavior associated with the item is using the "1" to "10" scale. In order to ensure that your most and least characteristic traits are identified, rate at least twenty (20) items "1" or "2", and no more than twenty (20) items at "9 or 10." You may find that putting 5 items in the "1" or "2" and 5 in the "9" or "10" ranges on EACH page is easiest way to accomplish this. Rate the remaining items within the range "3" and "8" depending upon your assessment of how characteristic of you each item is.

Less Characteristic	①②	③④⑤⑥⑦⑧	⑨⑩	**More Characteristic**

1B Shows a sincere interest in suggestions from members of the work group ①② ③④⑤⑥⑦⑧ ⑨⑩

APPENDIX C

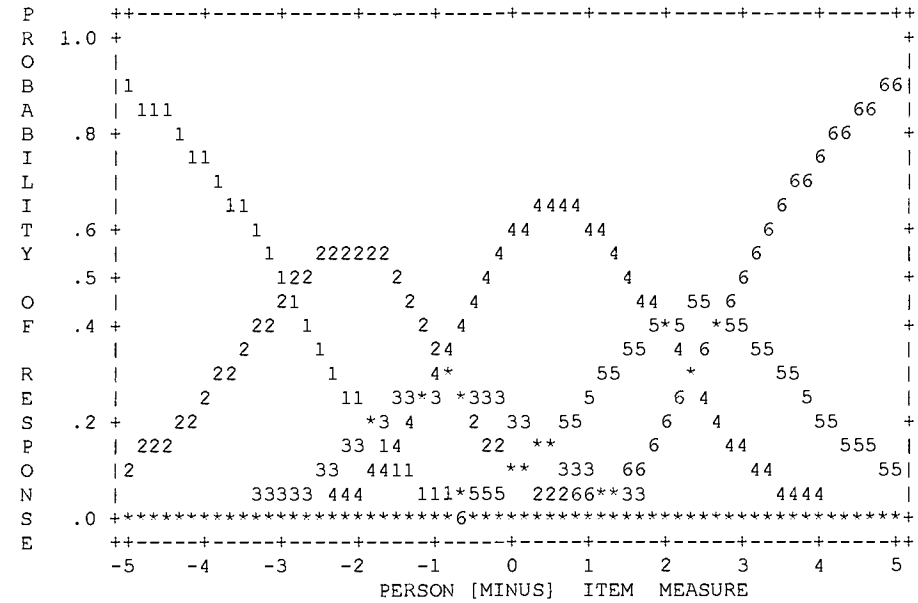
Items on the Group Involvement Scale.

ITEM	STATEMENT
1	Shows a sincere interest in suggestions from members of the work group.
2	Seeks out the special talents and abilities of others to contribute to the quality of the team's product.
3	Encourages people to speak up even when their opinions differ from the opinions of the majority.
4	Develops a feeling of unity and sharing among co-workers.
5	Publicly shares credit for success with those who contributed.
6	Actively promotes the involvement of people having a stake in the outcome of the project or task.

APPENDIX D

Non-forced Distribution Instructions (NFD)

Category probabilities.



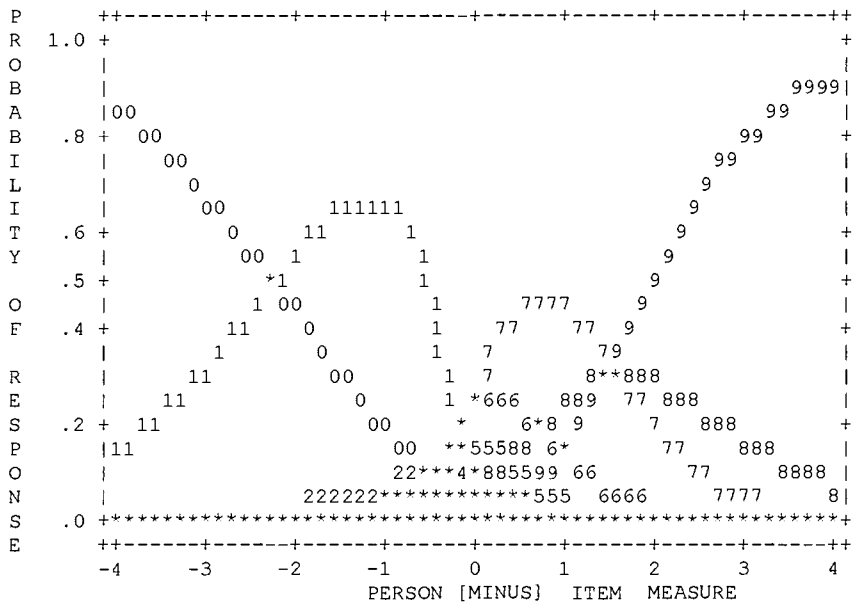
Summary of measured categories.

Category	Number of Observations	Step Calibration	Step Std. Error	Thurstone Threshold
1	7			
2	56	-2.91	.39	-3.01
3	82	-.71	.15	-1.20
4	324	-1.07*	.11	-.52
5	143	1.92	.10	1.69
6	72	2.77	.15	3.05

APPENDIX E

Forced Distribution Instructions (FD)

Category probabilities



Summary of measured categories.

Category	Number of Observations	Step Calibration	Step Std. Error	Thurstone Threshold
1	14			
2	102	-2.25	.27	-2.29
3	25	1.20	.11	-.43
4	21	.03*	.11	-.35
5	48	-.91*	.10	-.30
6	85	-.58	.09	-.21
7	137	-.40	.09	-.06
8	185	-.12	.09	.27
9	52	1.59	.13	1.26
10	21	1.44*	.23	2.00

AERA

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

TM032240

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: THE SUSCEPTIBILITY OF ITEM PARAMETERS TO INSTRUCTIONS FOR COMPLETION	
Author(s): JO-ANNE A. MCDONALD	
Corporate Source: UNIVERSITY OF DENVER	Publication Date:

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Level 1 ↑ ☑

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

Level 2A ↑ ☐

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 2B ↑ ☐

Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: JMcDonald	Printed Name/Position/Title: JO-ANNE MCDONALD, ABD	
Organization/Address: UNIV. OF DENVER, DENVER, CO 80208	Telephone: 303-221-6787	FAX:
	E-Mail Address: jmcdonal@du.edu	Date: 6-3-2000

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

ERIC CLEARINGHOUSE ON ASSESSMENT AND EVALUATION
UNIVERSITY OF MARYLAND
1129 SHRIVER LAB
COLLEGE PARK, MD 20772
ATTN: ACQUISITIONS

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
4483-A Forbes Boulevard
Lanham, Maryland 20706

Telephone: 301-552-4200
Toll Free: 800-799-3742
FAX: 301-552-4700
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com

EFF-088 (Rev. 2/2000)